Title

Gene Set Enrichment Analsyes Identiify Pathways Involved in Genetic Risk for Diabetic Retinopathy

Authors

Sobrin, Lucia
Susarla, Gayatri
Stanwyck, Lynn
et al.

Publication Date

2022

DOI

10.1016/j.ajo.2021.06.014

Peer reviewed

HHS Public Access

Author manuscript

Am J Ophthalmol. Author manuscript; available in PMC 2023 January 01.

Published in final edited form as:

Am J Ophthalmol. 2022 January ; 233: 111–123. doi:10.1016/j.ajo.2021.06.014.

Gene Set Enrichment Analyses Identify Pathways Involved in Genetic Risk for Diabetic Retinopathy

Lucia Sobrin, MD, MPH[1], **Gayatri Susarla, BS**[1], **Lynn Stanwyck, BS**[1], **John M. Rouhana, MA**[1], **Ashley Li, BS**[1], **Samuela Pollack, PhD**[2], **Robert P. Igo Jr., PhD**[3,*], **Richard A. Jensen, OD, PhD**[4], **Xiaohui Li, PhD**[5], **Maggie C.Y. Ng, PhD**[6,7,23], **Albert V. Smith, PhD**[8], **Jane Z. Kuo, MD, PhD**[9], **Kent D. Taylor, PhD**[5], **Barry I. Freedman, MD**[6,10], **Donald W. Bowden, PhD**[6,7], **Alan Penman, MBChB, PhD, MPH**[11], **Ching J. Chen, MD**[12], **Jamie E. Craig, MD, DPhil**[13], **Sharon G. Adler, MD**[14], **Emily Y. Chew, MD**[15], **Mary Frances Cotch, PhD**[15], **Brian Yaspan, PhD**[16], **Paul Mitchell, MBBS, MD, PhD**[17], **Jie Jin Wang, PhD**[17,18], **Barbara E.K. Klein, MD, MPH**[19], **Tien Y. Wong, MBBS, MMED, MPH, PhD**[18,20], **Jerome I. Rotter, MD**[5], **Kathyrn P. Burdon, PhD**[21], **Sudha K. Iyengar, PhD**[3], **Ayellet V. Segrè, PhD**[1,22]

[1]Department of Ophthalmology, Harvard Medical School, Massachusetts Eye and Ear Infirmary, Boston, MA

[2]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA

[3]Department of Population and Quantitative Health Sciences, Case Western University, Cleveland, OH

[4]Cardiovascular Health Research Unit, Department of Medicine, Epidemiology and Health Services, University of Washington, Seattle, WA

[5]Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA

[6]Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, NC

[7]Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC

[8]Department of Medicine, University of Iceland, Reykjavík, Iceland

Corresponding Author: Lucia Sobrin, MD, MPH, Massachusetts Eye and Ear Infirmary, Department of Ophthalmology, Harvard Medical School, 243 Charles Street, Boston, Massachusetts 02114, Tel: 617-573-3779, Lucia_Sobrin@meei.harvard.edu.

*Posthumous

C. Contributions of Authors

Design of the Study (LS, AVS); Conduct of the study (LS, GS, LKS, JH, AL, AVS); Collection, management, analysis and interpretation of the data (LS, GS, LKS, JH, AL, SP, RPI, RAJ, XL, MCYN, AVS, JZK, KDS, BIF, DWB, MFC, BY, JJW, BEKK, TYW, KPB, JIR, SKI, AVS); Preparation, review or approval of the manuscript (LS, GS, LKS, JH, AL, SP, RPI, RAJ, XL, MCYN, AVS, JZK, KDS, BIF, DWB, MFC, BY, JJW, BEKK, TYW, KPB, JIR, SKI, AVS)

B. Financial Disclosures

BLY is a full-time employee of Genentech Inc. and holds stock and stock options in Roche. JZK is employed by Sun Pharmaceutical Industries, Inc.; however, the current employer is not in any way involved in this study. All other authors declare no financial disclosures.

[9.]Medical Affairs, Ophthalmology, Sun Pharmaceutical Industries, Inc, Princeton, NJ

[10.]Department of Internal Medicine, Section on Nephrology, Wake Forest School of Medicine, Winston-Salem, NC

[11.]Department of Preventive Medicine, John D. Bower School of Population Health, University of Mississippi Medical Center, Jackson, MS

[12.]Department of Ophthalmology, University of Mississippi Medical Center, Jackson, MS

[13.]FHMRI Eye & Vision, Flinders University, Bedford Park SA, Australia

[14.]Department of Nephrology and Hypertension, Los Angeles Biomedical Research Institute at Harbor-University of California, Torrance, CA

[15.]Division of Epidemiology and Clinical Applications, National Eye Institute, National Institutes of Health, Bethesda, MD

[16.]Genentech Inc., South San Francisco, CA

[17.]Centre for Vision Research, Westmead Institute for Medical Research, University of Sydney, Sydney, Australia

[18.]Duke-NUS Medical School, Singapore

[19.]Department of Ophthalmology and Visual Sciences, University of Wisconsin-Madison, Madison, WI

[20.]Singapore Eye Research Institute, Singapore National Eye Centre, Singapore

[21.]Menzies Institute for Medical Research, University of Tasmania, Hobart, Australia

[22.]Broad Institute of Harvard and MIT, Cambridge, MA

[23.]Vanderbilt Genetics Institute and Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN

Abstract

Purpose: To identify functionally related genes associated with diabetic retinopathy (DR) risk using gene set enrichment analyses (GSEA) applied to genome-wide association study (GWAS) meta-analyses.

Methods: We analyzed DR GWAS meta-analyses performed on 3,246 Europeans and 2,611 African Americans with type 2 diabetes. Gene sets relevant to five key DR pathophysiology processes were investigated: tissue injury, vascular events, metabolic events and glial dysregulation, neuronal dysfunction, and inflammation. Keywords relevant to these processes were queried in four pathway and ontology databases. Two GSEA methods, Meta-Analysis Gene set Enrichment of variaNT Associations (MAGENTA) and Multi-marker Analysis of GenoMic Annotation (MAGMA) were used. Gene sets were defined to be enriched for gene associations with DR if the P value corrected for multiple testing (Pcorr) was <.05.

Results: Five gene sets were significantly enriched for multiple modest genetic associations with DR in one method (MAGENTA or MAGMA) and also at least nominally significant (uncorrected P <.05) in the other method. These pathways were regulation of the lipid catabolic process

(2-fold enrichment, Pcorr=.014); nitric oxide biosynthesis (1.92-fold enrichment, Pcorr=.022); lipid digestion, mobilization and transport (1.6-fold enrichment, P=.032); apoptosis (1.53-fold enrichment, P=.041); and retinal ganglion cell degeneration (2-fold enrichment, Pcorr=.049). The interferon gamma (*IFNG*) gene, previously implicated in DR by protein-protein interactions in our GWAS, was among the top ranked genes in the nitric oxide pathway (best variant P=.0001).

Conclusions: These GSEA indicate that variants in genes involved in oxidative stress, lipid transport and catabolism and cell degeneration are enriched for genes associated with DR risk.

INTRODUCTION

Diabetic retinopathy (DR) is a leading cause of blindness.[1] Established risk factors include longer duration of diabetes (DoD) and poor glycemic control.[2] Some populations, including African Americans, have been found to have a higher risk of developing DR compared with populations of European ancestry even after adjusting for these established risk factors.[3–7] Genetic factors are also implicated, with heritability of 52% for proliferative diabetic retinopathy (PDR).[8, 9] However, traditional individual candidate gene association studies have not been successful in identifying the underlying genetic architecture for DR. Furthermore, genome-wide association studies (GWAS) of DR to date have not had sufficient power to detect reproducible single DNA variants associations with the disease (tens to hundreds of thousands of individuals needed).[10–17]

Gene set enrichment analysis (GSEA) applied to GWAS variant data is a method that tests whether sets of functionally related genes are enriched for genetic associations with a polygenic disease or trait.[18] Previous studies have shown that GSEA of GWAS has the potential to detect associations likely missed by single-marker analysis.[19, 20] GSEA has been used successfully for various multifactorial diseases, such as type 2 diabetes and bipolar disorder, to determine if there is enrichment of genes in pathways implicated in disease pathogenesis among the top ranked genetic associations in GWAS.[19, 21–26]

We have previously collaborated to execute the largest DR GWAS to date.[17] The purpose of this study is to identify functionally related genes that affect risk of DR using GSEA on this GWAS dataset. We hypothesize that common variants associated with DR risk will affect genes that cluster in specific pathogenic pathways or biological processes and that both statistical and explanatory power can be gained by testing for enrichment of multiple modest genetic associations at the gene-set level, using GSEA, compared to testing genetic variants individually. This may be particularly beneficial for studies such as our latest DR GWAS studies, where only few variants passed genome-wide significance.

METHODS

All studies conformed to the Declaration of Helsinki tenets and were Health Insurance Portability and Accountability Act (HIPAA) compliant. Written informed consent was obtained from all participants. Institutional Review Board (IRB) approval was obtained prospectively by each individual study for collection of DNA and genotyping. The Massachusetts Eye and Ear Infirmary IRB approved the analysis of de-identified datasets from all the cohorts centrally at the Massachusetts Eye and Ear Infirmary.

GWAS META-ANALYSES ANALYZED

The study participants were the patients included in the discovery phase of a previously published DR GWAS.[17] These patients were from a consortium of 11 DR genetic studies which included 3,246 European and 2,611 African American patients. [11–13, 17, 27–30] All patients had type 2 diabetes which was defined as a fasting plasma glucose (FPG) ≥ 126 mg/dL or a hemoglobin A_{1C} (HbA_{1C}) ≥ 6.5% [31] with onset of the diabetes after age 30 years. Table 1 summarizes the DR phenotyping protocols and covariates by cohort. Phenotyping protocols have been previously described.[3, 9, 17, 32–40] All of these participants had genome-wide genotyping and were part of the GWAS. The GWAS analyses were performed with liability threshold (LT) modeling of DoD and glycemic control using LTSCORE,[41] and executed separately for the African American and European cohorts. Only variants on the autosomes were analyzed in these DR GWAS, hence genes on the sex chromosomes were not included in the GSEA analysis. We examined for any differences in the distribution of DR severity between men and women in the European and African American GWAS using a two-sided Wilcoxon rank sum test in each population.

The GSEA analyses for the present study used the GWAS meta-analyses summary statistics from the previous publication.[17] For this GSEA, we examined two DR case-control definitions with different Early Treatment Diabetic Retinopathy Study (ETDRS) score thresholds for cases and controls.[42] The first compared patients with PDR to those without PDR (Early treatment diabetic retinopathy study (ETDRS) ≥ 60 vs. ETDRS < 60, henceforth the PDR analysis). The second compared those with PDR to those without DR (ETDRS ≥ 60 vs. ETDRS < 14, henceforth the extremes of DR analysis). We chose to examine these two case-control definitions out of the total of four case-control definitions originally included the GWAS paper, because the individual variants with the most significant findings came from these two case-control definitions that have PDR as their case definition. This is consistent with the fact that PDR has a higher heritability than overall DR.[9] Table 1 shows the available samples by cohort and ETDRS score thresholds. Therefore, in total there were four GWAS meta-analyses datasets on which GSEA were run:

1. African Americans, PDR analysis
2. Europeans, PDR analysis
3. African Americans, Extremes of DR analysis
4. Europeans, Extremes of DR analysis

GENE SET ENRICHMENT ANALYSES

Extraction of gene sets—The gene sets that were examined in the GSEA were chosen based on their relevance to the pathophysiology of DR as summarized in Table 2 of a seminal paper on this subject.[43] The five pathophysiologic pathways from this table were tissue injury, vascular events, metabolic events and glial dysregulation, neuronal dysfunction, and inflammation. These pathways were broken down into keywords that were used to search in four gene set databases: Reactome Pathway Database (https://reactome.org), the Kyoto Encyclopedia of Genes and Genomes (KEGG, https://www.genome.jp/kegg), Gene Ontology (GO; https://geneontology.org), and Mouse Genome

Informatics (MGI; https://www.informatics,jax.org). Supplementary Table 1 lists all the pathophysiologic pathways and the resulting keywords that were queried and the respective gene sets that were identified from those searches. Each keyword was searched individually. Because some of the search terms were very general, the resultant gene sets were pruned by a clinician scientist with expertise in DR (LS) to include only those gene sets that truly reflect the pathophysiologic pathways in DR from the seminal paper.[43] We tested a total of 207 gene sets (143 GO, 13 KEGG, 41 MGI, and ten REACTOME gene sets).

GSEA of GWAS analysis—To increase the robustness of the results, the identified gene sets and the four DR GWAS datasets defined above were analyzed using two different GSEA methods: Meta-Analysis Gene set Enrichment of variaNT Associations (MAGENTA; http://www.broadinstitute.org/mpg/magenta) and Multi-marker Analysis of GenoMic Annotation (MAGMA; http://ctg.cncr.nl/software/magma). Both methods can be applied to GWAS summary statistics, leveraging the statistical power of large GWAS meta-analyses. MAGENTA is a rank and multivariate regression-based method that tests whether sets of functionally related genes (e.g., biological pathways) are enriched for highly ranked gene associations with a polygenic disease or trait more than would be expected by chance, correcting for confounding factors, including gene size and local linkage disequilibrium (LD)[19]. MAGMA is a gene set analysis tool that uses multiple linear regression models to assess whether genes in a given gene set are more strongly associated with a given polygenic trait compared to all other genes in the genome, correcting for confounding factors, such as LD between variants and gene size.[20] Both methods were applied to all genotyped and imputed variants in the four DR GWAS meta-analyses defined above.

For mapping of variant association P values to genes, we tested two gene boundary definitions: (1) –5 kilobases (kb) upstream and +5 kb downstream from the transcript start and end sites, respectively, to capture coding variants in the genes themselves and flanking regulatory regions, and (2) −110 kb upstream and +0 kb downstream from the transcript start and end sites, respectively, to capture additional potential regulatory causal variants in addition to coding variants. In both methods, all genes in the genome were scored based on the most significant association p-value of all variants within each gene's window using the two boundary definitions.

In the MAGENTA analysis, stepwise multivariate linear regression analysis is used to correct for confounding effects on assigning the most significant variant association P value per gene, including gene size, local variant density, and local LD. The LD covariate was computed as the number of LD-independent variants (r^2>0.5) per gene region, using the African American and European subpopulations in 1000 Genomes Project Phase 3 for the corresponding ancestral backgrounds in the DR GWAS. The adjusted gene association P values were subsequently used to rank genes in the genome with respect to their likelihood of association with DR, and permutation analysis was used to compute a gene set enrichment P value for each gene set of interest. The gene set enrichment P value calculated by MAGENTA assesses the overrepresentation of highly ranked gene association P values above a given enrichment cutoff, compared to multiple randomly sampled gene sets from the genome with equal gene set size. Physical proximity along the chromosome between two or more genes in a given gene set was corrected for by collapsing all genes that share the same

most significant variant to one effective gene, retaining the gene with the most significant adjusted gene association P value. The human leukocyte antigen (HLA) region was removed due to high LD and gene density in the region, making it difficult to disentangle the putative causal gene if an association signal exists in the region. The 75th and 95th percentiles of all adjusted gene P values were used as the enrichment cutoffs.

In the MAGMA analysis, multiple linear principal components regression analysis is used to correct for LD between variants in scoring genes based on the most significant variant P values. The estimates of LD between variants in gene regions were also computed using Phase 3 of the 1000 Genomes Project and the African American and European subsets for the corresponding DR GWAS. The gene P value results from the analyses were converted to Z-values that were inputted into the GSEA. A generalized linear regression model of gene Z-values was used to assess whether genes in a given gene set are more strongly associated with a given polygenic trait than all other genes in the genome, correcting for gene size, gene density and differences in underlying GWAS sample size in the meta-analysis by adding these variables as covariates in the gene or gene set level models.[20] For this study, we chose the competitive gene set analysis option in MAGMA, similar what is done in MAGENTA.

For both MAGENTA and MAGMA, only gene sets with ten to 2000 genes were included in the analysis because very small or large gene sets are subject to unstable results from violation of some of the assumptions of these GSEA methods. To address the issue of multiple hypothesis testing, Benjamini-Hochberg (BH) correction of the gene set enrichment P value was carried out for both methods.[44] To compute the BH-adjusted P values, all gene sets were ranked for a given GWAS-ancestry-window-database group based on the gene sets' uncorrected GSEA P value in ascending order. Then the uncorrected GSEA P value was multiplied by the total number of gene sets tested for the given database and divided by the rank of each specific gene set. The BH correction is more appropriate than a Bonferroni correction given the overlap between gene sets tests. A BH-corrected P value < .05 was considered statistically significant. We prioritized gene sets with BH-corrected P value < .05 with at least one method and uncorrected P value < .05 with the other method.

Clustering of significant gene sets based on gene membership similarity.—To assess the number of functional modules represented by the gene sets with significant enrichment for DR associations, we clustered the 15 significant gene sets reported in Table 2 based on fraction of genes that overlap between all pairwise gene set comparisons, using hierarchical clustering and Euclidean distance. We performed the clustering considering either all genes in the gene sets (Supplementary Figure 6) or just the leading edge genes in each gene set (top ranked DR-associated genes above the 75th percentile enrichment cutoff based on MAGENTA gene P values) in each gene set (Figure 2).

Testing for sex-biased expression among leading edge genes in significant gene sets—Differentially expressed genes between females and males were taken from a study that inspected the effect of sex on gene expression in 44 GTEx tissues (Release v8), including DR-relevant tissues: tibial artery and 11 brain regions.[45] Sex-biased genes were computed with a multivariate adaptive shrinkage (MASH) method and were considered

significant at a local false sign rate (LFSR) ≤ 0.05, correcting for multiple hypothesis testing.[46] We assessed the enrichment of sex-biased gene expression among leading edge genes in the significant gene sets compared to non-leading edge genes using a two-sided Fisher's exact test, and the enrichment of sex-biased genes among the leading edge genes compared to the observed number of sex-biased genes amongst all genes expressed in the given tissue and given the gene set size, using the hypergeometric cumulative distribution function.

RESULTS

The GWAS meta-analyses for the GSEA included 1,097 African American and 398 European PDR cases (ETDRS ≥ 60). For the PDR analysis, they were compared to 1,514 African and 2,848 European controls without PDR (ETDRS < 60), respectively. For the Extremes of DR analysis, they were compared to 941 African American and 1,970 European controls without DR (ETDRS < 14), respectively. There was no difference in the distribution of ETDRS severity between men and women in the African American (P=.47) and European populations (P=0.99) in this GWAS (Supplementary Figure 1).

Out of 207 gene sets tested, 15 gene sets were found to be significant in either MAGENTA or MAGMA analyses after BH correction (Table 2; full results in Supplementary Table 2). No gene set was significant in both MAGENTA and MAGMA after multiple hypothesis correction. The gene sets most significant with MAGENTA were based on the African American GWAS, while the gene sets most significant with MAGMA were based on the European GWAS. The distribution of gene association P values based on the European DR GWAS showed slightly higher excess of low gene P values compared to the African DR GWAS with both MAGENTA and MAGMA (Supplementary Figures 2 and 3), and the overall correlation of all gene P values between the two methods within the same ancestral GWAS was high (Spearman's rho=0.87–0.91, $P<10^{-70}$; Supplementary Figure 4), but not between ancestries (Spearman's rho=0.008–0.13; Supplementary Figure 5).

There were five gene sets that were significant by BH correction in one method and had an uncorrected P value < .05 in the other method. Figure 1 shows the gene P value distributions for these five gene sets. Enrichment of genes with low P values was most pronounced for the regulation of nitric oxide biosynthetic process, and regulation of lipid catabolic process gene sets (Figure 1A and 1C). The Pearson and Spearman's correlation coefficients comparing the gene P values from MAGENTA and MAGMA showed high correlation for all five gene sets (Spearman's rho=0.86–0.93, $P<10^{-6}$, Table 3). Supplementary Table 3 lists the individual genes within these five gene sets ranked based on their gene-level DR association P values from MAGENTA and MAGMA, as well as the P value of the most associated variant within or around each gene in the given GWAS.

The regulation of nitric oxide biosynthetic process gene set in GO (African American PDR analysis) was significant after BH correction in MAGENTA and had an uncorrected P value of = .0026 in MAGMA. Among the 46 genes examined in MAGENTA in the nitric oxide biosynthetic process gene set, 23 genes were above the enrichment cut-off (listed in Table 3), where only 12 where expected by chance, yielding a 1.92-fold enrichment for this

gene set, which is among the highest in this analysis (Table 2). This analysis suggests that eleven genes among the top 23 genes ranked based on their DR gene P values (leading edge genes) are likely to be true DR associations, even though none of the top variants for each of these genes passed genome-wide significance in the current GWAS. Larger GWAS meta-analyses will be needed to replicate these associations. In Supplementary Table 3, the 23 leading edge genes (the genes above the enrichment cutoff) are listed. Among the top genes in this pathway is interferon gamma (*IFNG*), which was also found to be enriched for protein-protein interactions in our previous DR GWAS analyses.[17]

Two other gene sets were significant after BH correction in MAGENTA, and significant before correction in MAGMA: regulation of lipid catabolic process in GO (African Americans, PDR analysis) and apoptosis in KEGG (African Americans, extremes of DR analysis). These gene sets had a fold-enrichment of 2 and 1.53, respectively, in MAGENTA. *CAPN2*, a gene for a calcium-activated neutral protease, is one of the genes in the KEGG apoptosis pathway and it also has an expression quantitative trait locus (eQTL) that was implicated by the top finding from our original GWAS in the extremes of DR analysis, variant rs4121487 in the nuclear VCP-like (*NVL)* gene.[17]

The two gene sets that were significant in MAGMA after BH correction and had a P value < .05 in MAGENTA were lipid digestion mobilization and transport in Reactome (European, PDR analysis) and retinal ganglion cell degeneration in MGI (European, PDR analysis). These gene sets had a fold-enrichment of 1.6 and 2.0, respectively, in MAGENTA.

There were also two gene sets related to vascular endothelial growth factor (VEGF) which were significant in either MAGMA or MAGENTA (Table 2). The KEGG VEGF signaling pathway was significant in the MAGENTA analysis (African Americans, extremes of DR analysis) and the VEGF receptor signaling pathway was significant in the MAGMA analysis (European, PDR analysis)

Clustering of the 15 significant gene sets in Table 2 based on their leading edge genes computed with MAGENTA, suggests 8 key biological processes that might affect DR risk (Figure 2). These include (in order of clustering): (1) lipid transport, (2) retinal degeneration and tight junction interactions, (3) platelet derived growth factor receptor signaling, (4) apoptosis and VEGF signaling (KEGG), (5) nitric oxide biosynthesis and tissue development, (6) lipid and lipoprotein metabolism and regulation, (7) post-translational protein modification, and (8) VEGF receptor signaling (GO). Since the leading edge genes are determined by the GWAS variant P values, the gene sets clustered first by population of the GWAS and then by gene set type, compared to clustering of gene sets considering all genes (Supplementary Figure 6).

Lastly, given the suggested effect of sex on DR susceptibility, we examined whether the leading edge genes (enriched for DR-associated genes) in the five significant gene sets were enriched for differential gene expression between females and males in DR-relevant tissues (blood vessel and brain tissue).[47–49] We found that 6–24% of the leading edge genes across the five gene sets showed sex-biased expression in tibial artery and 13–35% in eleven different brain regions in GTEx (Supplementary Table 4).[45] However, these fractions were

not significantly higher compared to the non-leading edge genes in each gene set (Fisher's exact test P>0.28; Supplementary Table 5), not were the leading edge genes significantly enriched for sex-biased genes compared to what would be expected by chance given the gene set size and number of sex-biased genes among all genes expressed in each tissue (Hypergeometric P>0.127; Supplementary Table 6).

DISCUSSION

This GSEA applied to summary statistics data from a GWAS for DR provides evidence that biological processes in five pathways with prior evidence for involvement in DR are enriched for multiple modest genetic associations with DR. These pathways are nitric oxide biosynthesis; regulation of the lipid catabolic process; lipid digestion, mobilization and transport; apoptosis; and retinal ganglion cell degeneration. This supports a causal contribution to DR risk for genes in these pathways.

There is extensive evidence linking nitric oxide overexpression and DR.[50] One interesting finding among the genes in the nitric oxide biosynthesis gene set is the highly ranked *IFNG* gene. In a previous analysis examining significantly enriched protein networks among loci with the highest statistical significance for association with DR, we identified a significant protein network that also included *IFNG*, and this was also within the PDR Analysis in African Americans.[17] Interferon-gamma is highly expressed in ocular tissues from PDR patients and polymorphisms within this gene have been previously associated with PDR.[51] In this previous study, rs2430561 was the variant associated with PDR, and that variant is in modest LD [r^2=0.256, D'=0.99 based on European subset in GTEx release v8[52]and r^2=0.026, D'=1 based on African American subset in GTEx release v8)] with the top variant from this analysis, rs2069733, which is 2.29 kb upstream of rs2430561.

Two lipid pathways were also significant in this analysis: lipid catabolism and lipid transport. Both were identified in the PDR analysis, one in Europeans and the other in African Americans. Dyslipidemia has also been extensively associated with DR.[53] Among the genes in the leading edge of the lipid digestion, mobilization and transport gene set is *APOA1*; plasma levels of Apo A1 have been inversely correlated with DR severity.[53] *APOA2* is the one gene that is part of the leading edge for both of these lipid-related gene sets.

With regards to apoptosis and retinal ganglion cell degeneration, pericyte apoptosis is one of the earliest events in the development of DR,[54] and growing evidence indicates that degeneration of retinal ganglion cells also occurs before clinical signs of DR.[55] We note that *CAPN2,* the gene targeted by an eQTL of the top genome-wide significant findings from our GWAS in European ancestry individuals, was one of the genes within the KEGG apoptosis gene set. Although it was not a leading edge gene, possibly because the DR-associated eQTL lies outside the gene boundaries used in the GSEA (532 kb downstream of *CAPN2* transcription start site), it was close to the enrichment cut-off in the gene set, and the finding adds some support for a role for this gene in DR genetic risk.

The strengths of this study include the use of two different gene set analysis methods, MAGENTA and MAGMA, to evaluate the contribution of gene sets to DR risk. The two methods differ in their approaches, but despite these differences, there was common support from both methods for a role of lipid metabolism or transport, cell death, and VEGF signaling, and excellent concordance in the gene association scores between the methods, as evidenced by the Pearson and Spearman's correlations being very strong. The correlations were slightly less strong, but still clearly significant, for the retinal ganglion cell degeneration gene set, in part because the size of this gene set was significantly smaller than the others.

The observation that MAGENTA primarily found significant genes sets in the African American DR GWAS and MAGMA in the European DR GWAS is likely due to differences in the gene set enrichment statistical tests, as the adjusted gene level association P values computed by each of the methods highly correlate (Supplementary Figure 4). The rank-based approach in MAGENTA using the 75th percentile enrichment cut-off identifies enrichment of multiple weak effects in a given gene set,[19] while the regression-based approach of MAGMA identifies gene sets with an overall stronger average association with disease risk compared to all other genes in the genome. This is in concordance with the European DR GWAS displaying a slightly higher excess of low gene P values (stronger effects) than the African American DR GWAS (Supplementary Figure 3).

Currently none of the enriched gene sets found are significant after multiple hypothesis correction in both ancestral populations. This is consistent with the top-ranked associated genes being different between the European and African American GWAS (Supplementary Figure 5), which may be due to the limited power of the GWAS, especially for the European GWAS which had fewer PDR cases. It is also possible that there are environmental elements that may differentially influence the expression of genes in the two populations and account for differences in which genes are associated with DR in the two populations. The VEGF signaling pathway though represented by gene sets from two different databases was found to be significant in both ancestries.

There are some limitations to this study. First, our GWAS for DR is of modest size which may have limited our power to detect gene sets that were significant after correction with both methods, MAGENTA and MAGMA.[18] Based on simulations in MAGENTA to assess the power of its GSEA algorithm to detect enrichment of genes with modest effect sizes that would be missed with individual SNP analyses, we have >95% power to detect significant enrichment if >4% or >20% of genes are modestly associated with DR in gene sets with 25 or 100 genes, respectively,[19] which is in concordance with our results (Supplementary Table 2). We note that we did not correct for the additional multiple testing related to testing two gene boundaries and two different case-control definitions as there is significant overlap between these conditions and these are not independent tests; this approach is common in the field.[56–58] The two different populations, however, are independent. If we additionally would have corrected for these two populations in the BH correction, the strongest associations in MAGMA and MAGENTA still remain, and the weaker ones would become insignificant. Second, the gene boundary definitions we examined (+/− 5 kb and −110 kb upstream/+ 40kb downstream from the transcript start and end sites, respectively)

might miss causal variants if they are in regulatory regions outside of these boundaries, as regulatory variants may be hundreds of kilobases away from the target gene (e.g., 95% of eQTLs lie within 643 kb of the target gene's transcription start site).[52] We did not extend these boundaries further as we risked capturing and confusing neighboring gene signals with a larger boundary size. Still, our boundary sizes captured a significant amount of genetic variation. Finally, the genes in each gene set are not all equally relevant to different tissues. Some genes in a gene set may not be highly expressed in a particular tissue such as retina or retinal vasculature, and therefore may be less relevant for the DR phenotype. As tissue-specific gene sets are developed, it will be possible to refine the GSEA. Other limitations include the lack of longitudinal data that would allow examination of the influence of these gene sets on DR progression and the lack of data on therapies for diabetes and DR that would allow examination of the influence of these covariates on genetic risk for DR.

To our best knowledge, this is the first GSEA applied to genetic associations with DR. Within known pathophysiologically-relevant gene sets, the results of these analyses help us to rank which genes in these pathways are more likely implicated in genetic risk for DR and prioritize which genes and variants should be further investigated in additional studies (Supplementary Table 3). We find modest evidence for enrichment of variants involved in oxidative stress, lipid transport and catabolism and cell degeneration. Much larger GWAS datasets and datasets that include other populations that have higher prevalence of DR, such as the South Indian population, are needed to confirm and expand on our findings here.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

A. Funding/Support

We gratefully acknowledge support from the following organizations for this research: Research to Prevent Blindness, Inc., New York; National Eye Institute (EY16335; EY22302; EY11753; R01 EY023644; Core Grant EY001792); Massachusetts Lions Eye Research Fund; Alcon Research Institute; American Diabetes Association (1–11-CT-51); Harvard Catalyst.

The Age, Gene, Environment, Susceptibility - Reykjavik Study (AGES) was supported by the U.S. National Institutes of Health (NIH) through the Intramural Research Program of the National Institute of Aging (ZIAAG007380) and the National Eye Institute (ZIAEY00401), NIH contract number N01-AG-1–2100, Hjartavernd (the Icelandic Heart Association), the Althingi (Icelandic Parliament), and the University of Iceland Research Fund. We are indebted to the staff at the Icelandic Heart Association and to the AGES participants who volunteered their time and allowed us to contribute their data to this international project. The funders had no role in collection, management, analysis or interpretation of data nor were funders involved in the preparation, writing, or approval of the article, or the decision to submit the article for publication.

The Atherosclerosis Risk in Communities (ARIC) Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C), R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. The ARIC study has been funded in whole or in part with Federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services (contract numbers HHSN268201700001I, HHSN268201700002I, HHSN268201700003I, HHSN268201700004I and HHSN268201700005I). The authors

thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research. Funding support for “Building on GWAS for NHLBI-diseases: the U.S. CHARGE consortium” was provided by the NIH through the American Recovery and Reinvestment Act of 2009 (ARRA) (5RC2HL102419).

The Australian Genetics of Diabetic Retinopathy Study was supported by the National Health and Medical Research Council (NHMRC) of Australia [no. 595918] and the Ophthalmic Research Institute of Australia. K.P.B is supported by a Senior Research Fellowship from the NHMRC and J.E.C by a Practitioner Fellowship from the NHMRC.

The Blue Mountains Eye Study (BMES) was supported by the Australian National Health & Medical Research Council (NHMRC), Canberra Australia (NHMRC project grant IDs 974159, 211069, 302068, and Centre for Clinical Research Excellence in Translational Clinical Research in Eye Diseases, CCRE in TCR-Eye, grant ID 529923). The BMES GWAS and genotyping costs was supported by Australian NHMRC, Canberra Australia (NHMRC project grant IDs 512423, 475604 and 529912), and the Wellcome Trust, UK as part of Wellcome Trust Case Control Consortium 2 (IDs 085475/B/08/Z and 085475/08/Z).

The Cardiovascular Health Study was supported by contracts HHSN268201200036C, HHSN268200800007C, HHSN268201800001C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086, N01HC75150, and grants U01HL080295 and U01HL130114 from the National Heart, Lung, and Blood Institute (NHLBI), with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided by R01AG023629 from the National Institute on Aging (NIA). A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Subjects included in the present analysis consented to the use of their genetic information.

The Jackson Heart Study (JHS) is supported and conducted in collaboration with Jackson State University (HHSN268201300049C and HHSN268201300050C), Tougaloo College (HHSN268201300048C), and the University of Mississippi Medical Center (HHSN268201300046C and HHSN268201300047C) contracts from the National Heart, Lung, and Blood Institute (NHLBI) and the National Institute for Minority Health and Health Disparities (NIMHD). The authors also wish to thank the staffs and participants of the JHS. The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; the National Institutes of Health; or the U.S. Department of Health and Human Services.

MESA and the MESA SHARe projects are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts 75N92020D00001, HHSN268201500003I, N01-HC-95159, 75N92020D00005, N01-HC-95160, 75N92020D00002, N01-HC-95161, 75N92020D00003, N01-HC-95162, 75N92020D00006, N01-HC-95163, 75N92020D00004, N01-HC-95164, 75N92020D00007, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, UL1-TR-001420, UL1-TR-001881, and DK063491. Additional funding provided by the Intramural Research Program of the National Eye Institute (ZIAEY000403). Funding for SHARe genotyping was provided by NHLBI Contract N02-HL-64278. Genotyping was performed at Affymetrix (Santa Clara, California, USA) and the Broad Institute of Harvard and MIT (Boston, Massachusetts, USA) using the Affymetrix Genome-Wide Human SNP Array 6.0. The authors thank the other investigators, the staff, and the participants of the MESA study for their valuable contributions. A full list of participating MESA investigators and institutions can be found at http://www.mesa-nhlbi.org.

For Wake Forest School of Medicine study (WFU), genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSC268200782096C. This work was supported by National Institutes of Health grants R01 DK087914, R01 DK066358, R01 DK053591, DK070941, DK084149 and by the Wake Forest School of Medicine grant M01 RR07122 and Venture Fund.

This study was supported by the National Eye Institute of the National Institutes of Health (EY014684 to J.I.R.) and ARRA Supplement (EY014684–03S1, −04S1), the National Institute of Diabetes and Digestive and Kidney Disease grant DK063491 to the Southern California Diabetes Endocrinology Research Center, and the Eye Birth Defects Foundation Inc. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant, UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Research Center.

The FIND study was supported by grants U01DK57292, U01DK57329, U01DK057300, U01DK057298, U01DK057249, U01DK57295, U01DK070657, U01DK057303, and U01DK57304 from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) and, in part, by the Intramural Research Program of the NIDDK. Support was also received from the National Heart, Lung and Blood Institute grants U01HL065520,

U01HL041654, and U01HL041652. This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health (NIH), under contract N01-CO-12400 and the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research. This work was also supported by the National Center for Research Resources for the General Clinical Research Center grants: Case Western Reserve University, M01-RR-000080; Wake Forest University, M01-RR-07122; Harbor-University of California, Los Angeles Medical Center, M01-RR-00425; College of Medicine, University of California, Irvine, M01-RR-00827–29; University of New Mexico, HSC M01-RR-00997; and Frederic C. Bartter, M01-RR-01346. Computing resources were provided, in part, by the Wake Forest School of Medicine Center for Public Health Genomics. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

D. Other Acknowledgment: None

References

1. Yau JW, Rogers SL, Kawasaki R, et al. Global prevalence and major risk factors of diabetic retinopathy. Diabetes Care 2012;35(3):556–64. [PubMed: 22301125]
2. Klein R, Klein BE, Moss SE, Davis MD, DeMets DL. The Wisconsin epidemiologic study of diabetic retinopathy. III. Prevalence and risk of diabetic retinopathy when age at diagnosis is 30 or more years. Arch Ophthalmol 1984;102(4):527–32. [PubMed: 6367725]
3. Klein R, Marino EK, Kuller LH, et al. The relation of atherosclerotic cardiovascular disease to retinopathy in people with diabetes in the Cardiovascular Health Study. Br J Ophthalmol 2002;86(1):84–90. [PubMed: 11801510]
4. Leske C, Wu S, Hennis A, et al. Nine-year incidence of diabetic retinopathy in the Barbados Eye Studies. Arch Ophthalmol 2006;124(2):250. [PubMed: 16476895]
5. Emanuele N, Sacks J, Klein R, et al. Ethnicity, race, and baseline retinopathy correlates in the veterans affairs diabetes trial. Diabetes Care 2005;28(8):1954–1958. [PubMed: 16043738]
6. Harris EL, Sherman SH, Georgopoulos A. Black-white differences in risk of developing retinopathy among individuals with type 2 diabetes. Diabetes Care 1999;22(5):779–783. [PubMed: 10332681]
7. Emanuele N, Moritz T, Klein R, et al. Ethnicity, race, and clinically significant macular edema in the Veterans Affairs Diabetes Trial (VADT). Diabetes Res Clin Pract 2009;86(2):104–110. [PubMed: 19720420]
8. Looker HC, Nelson RG, Chew E, et al. Genome-wide linkage analyses to identify Loci for diabetic retinopathy. Diabetes 2007;56(4):1160–6. [PubMed: 17395753]
9. Arar NH, Freedman BI, Adler SG, et al. Heritability of the severity of diabetic retinopathy: the FIND-Eye study. Invest Ophthalmol Vis Sci 2008;49(9):3839–45. [PubMed: 18765632]
10. Sheu WH, Kuo JZ, Lee IT, et al. Genome-wide association study in a Chinese population with diabetic retinopathy. Hum Mol Genet 2013;22(15):3165–3173. [PubMed: 23562823]
11. Awata T, Yamashita H, Kurihara S, et al. A genome-wide association study for diabetic retinopathy in a Japanese population: potential association with a long intergenic non-coding RNA. PloS one 2014;9(11):e111715. [PubMed: 25364816]
12. Fu YP, Hallman DM, Gonzalez VH, et al. Identification of Diabetic Retinopathy Genes through a Genome-Wide Association Study among Mexican-Americans from Starr County, Texas. J Ophthalmol 2010; 2010.
13. Grassi MA, Tikhomirov A, Ramalingam S, Below JE, Cox NJ, Nicolae DL. Genome-wide meta-analysis for severe diabetic retinopathy. Hum Mol Genet 2011;20(12):2472–81. [PubMed: 21441570]
14. Huang YC, Lin JM, Lin HJ, et al. Genome-wide association study of diabetic retinopathy in a Taiwanese population. Ophthalmology 2011;118(4):642–8. [PubMed: 21310492]
15. Burdon KP, Fogarty RD, Shen W, et al. Genome-wide association study for sight-threatening diabetic retinopathy reveals association with genetic variation near the GRB2 gene. Diabetologia 2015;58(10):2288–97. [PubMed: 26188370]
16. Shtir C, Aldahmesh MA, Al-Dahmash S, et al. Exome-based case-control association study using extreme phenotype design reveals novel candidates with protective effect in diabetic retinopathy. Hum Genet 2016;135(2):193–200. [PubMed: 26693933]

17. Pollack S, Igo R P Jr., Jensen RA, et al. Multiethnic Genome-Wide Association Study of Diabetic Retinopathy Using Liability Threshold Modeling of Duration of Diabetes and Glycemic Control. Diabetes 2019;68(2):441–456. [PubMed: 30487263]
18. de Leeuw CA, Neale BM, Heskes T, Posthuma D. The statistical properties of gene-set analysis. Nat Rev Genet 2016;17(6):353–64. [PubMed: 27070863]
19. Segre AV, Consortium D, investigators M, et al. Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. PLoS Genet 2010;6(8).
20. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol 2015;11(4):e1004219. [PubMed: 25885710]
21. Voight BF, Scott LJ, Steinthorsdottir V, et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet 2010;42(7):579–89. [PubMed: 20581827]
22. Morris AP, Voight BF, Teslovich TM, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet 2012;44(9):981–90. [PubMed: 22885922]
23. Segre AV, Wei N, Consortium D, Investigators M, Altshuler D, Florez JC. Pathways targeted by antidiabetes drugs are enriched for multiple genes associated with type 2 diabetes risk. Diabetes 2015;64(4):1470–83. [PubMed: 25368101]
24. Lango Allen H, Estrada K, Lettre G, et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature 2010;467(7317):832–8. [PubMed: 20881960]
25. Stahl EA, Breen G, Forstner AJ, et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. Nat Genet 2019;51(5):793–803. [PubMed: 31043756]
26. Khawaja AP, Cooke Bailey JN, Wareham NJ, et al. Genome-wide analyses identify 68 new loci associated with intraocular pressure and improve risk prediction for primary open-angle glaucoma. Nat Genet 2018;50(6):778–782. [PubMed: 29785010]
27. Hosseini SM, Boright AP, Sun L, et al. The association of previously reported polymorphisms for microvascular complications in a meta-analysis of diabetic retinopathy. Human genetics 2015;134(2):247–57. [PubMed: 25487307]
28. Grassi MA, Mazzulla DA, Knudtson MD, et al. Patient self-report of prior laser treatment reliably indicates presence of severe diabetic retinopathy. Am J Ophthalmol 2009;147(3):501–4. [PubMed: 19054495]
29. Grassi MA, Tikhomirov A, Ramalingam S, et al. Replication analysis for severe diabetic retinopathy. Invest Ophthalmol Vis Sci 2012;53(4):2377–81. [PubMed: 22427569]
30. Meng W, Shah KP, Pollack S, et al. A genome-wide association study suggests new evidence for an association of the NADPH Oxidase 4 (NOX4) gene with severe diabetic retinopathy in type 2 diabetes. Acta Ophthalmol 2018.
31. Report of the expert committee on the diagnosis and classification of diabetes mellitus. Diabetes Care 2003;26 Suppl 1:S5–20. [PubMed: 12502614]
32. Sobrin L, Green T, Sim X, et al. Candidate gene association study for diabetic retinopathy in persons with type 2 diabetes: the Candidate gene Association Resource (CARe). Invest Ophthalmol Vis Sci 2011;52(10):7593–602. [PubMed: 21873659]
33. Wong TY, Klein R, Islam FM, et al. Diabetic retinopathy in a multi-ethnic cohort in the United States. Am J Ophthalmol 2006;141(3):446–455. [PubMed: 16490489]
34. Klein R, Sharrett AR, Klein BE, et al. The association of atherosclerosis, vascular risk factors, and retinopathy in adults with diabetes: the Atherosclerosis Risk in Communities Study. Ophthalmology 2002;109(7):1225–34. [PubMed: 12093643]
35. Kaidonis G, Abhary S, Daniell M, et al. Genetic study of diabetic retinopathy: recruitment methodology and analysis of baseline characteristics. Clinical & experimental ophthalmology 2014;42(5):486–93. [PubMed: 24112246]
36. Mitchell P, Smith W, Wang JJ, Attebo K. Prevalence of diabetic retinopathy in an older community. The Blue Mountains Eye Study. Ophthalmology 1998;105(3):406–11. [PubMed: 9499768]
37. Gunnlaugsdottir E, Halldorsdottir S, Klein R, et al. Retinopathy in old persons with and without diabetes mellitus: the Age, Gene/Environment Susceptibility--Reykjavik Study (AGES-R). Diabetologia 2012;55(3):671–80. [PubMed: 22134840]

38. Nguyen QD, Brown DM, Marcus DM, et al. Ranibizumab for diabetic macular edema: results from 2 phase III randomized trials: RISE and RIDE. Ophthalmology 2012;119(4):789–801. [PubMed: 22330964]
39. Penman A, Hoadley S, Wilson JG, Taylor HA, Chen CJ, Sobrin L. P-selectin Plasma Levels and Genetic Variant Associated With Diabetic Retinopathy in African Americans. Am J Ophthalmol 2015;159(6):1152–1160 e2. [PubMed: 25794792]
40. Kuo JZ, Guo X, Klein R, et al. Systemic soluble tumor necrosis factor receptors 1 and 2 are associated with severity of diabetic retinopathy in Hispanics. Ophthalmology 2012;119(5):1041–6. [PubMed: 22330960]
41. Zaitlen N, Lindstrom S, Pasaniuc B, et al. Informed conditioning on clinical covariates increases power in case-control association studies. PLoS genetics 2012;8(11):e1003032. [PubMed: 23144628]
42. Grading diabetic retinopathy from stereoscopic color fundus photographs--an extension of the modified Airlie House classification. ETDRS report number 10. Early Treatment Diabetic Retinopathy Study Research Group. Ophthalmology 1991;98(5 Suppl):786–806. [PubMed: 2062513]
43. Antonetti DA, Klein R, Gardner TW. Diabetic retinopathy. N Engl J Med 2012;366(13):1227–39. [PubMed: 22455417]
44. Chen SY, Feng Z, Yi X. A general introduction to adjustment for multiple comparisons. J Thorac Dis 2017;9(6):1725–1729. [PubMed: 28740688]
45. Oliva M, Munoz-Aguirre M, Kim-Hellmuth S, et al. The impact of sex on gene expression across human tissues. Science 2020;369(6509).
46. Urbut S, Wang G, Carbonetto P, Stephens M. Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. Nat Genet 2019;51(1):187–195. [PubMed: 30478440]
47. Ozawa G, Bearse M Jr, Adams A. Male-female differences in diabetic retinopathy? Curr Eye Res 2015;40(2):234–246. [PubMed: 25545999]
48. Seghieri G, Policardo L, Anichini R, et al. The Effect of Sex and Gender on Diabetic Complications. Curr Diabetes Rev 2017;13(2):148–160. [PubMed: 27183843]
49. Cherchi S, Gigante A, Spanu M, et al. Sex-Gender Differences in Diabetic Retinopathy. Diabetology 2020;1(1):1–10.
50. Zheng L, Kern TS. Role of nitric oxide, superoxide, peroxynitrite and PARP in diabetic retinopathy. Front Biosci (Landmark Ed) 2009;14:3974–87. [PubMed: 19273327]
51. Paine SK, Basu A, Mondal LK, et al. Association of vascular endothelial growth factor, transforming growth factor beta, and interferon gamma gene polymorphisms with proliferative diabetic retinopathy in patients with type 2 diabetes. Mol Vis 2012;18:2749–57. [PubMed: 23213275]
52. Consortium GT. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science 2020;369(6509):1318–1330. [PubMed: 32913098]
53. Chang YC, Wu WC. Dyslipidemia and diabetic retinopathy. Rev Diabet Stud 2013;10(2–3):121–32. [PubMed: 24380088]
54. Beltramo E, Porta M. Pericyte loss in diabetic retinopathy: mechanisms and consequences. Curr Med Chem 2013;20(26):3218–25. [PubMed: 23745544]
55. Catalani E, Cervia D. Diabetic retinopathy: a matter of retinal ganglion cell homeostasis. Neural Regen Res 2020;15(7):1253–1254. [PubMed: 31960808]
56. Giri A, Hellwege JN, Keaton JM, et al. Trans-ethnic association study of blood pressure determinants in over 750,000 individuals. Nat Genet 2019;51(1):51–62. [PubMed: 30578418]
57. Shah RL, Guggenheim JA, Eye UKB, Vision C. Genome-wide association studies for corneal and refractive astigmatism in UK Biobank demonstrate a shared role for myopia susceptibility loci. Hum Genet 2018;137(11–12):881–896. [PubMed: 30306274]
58. Liu H, Sun Y, Zhang X, et al. Integrated Analysis of Summary Statistics to Identify Pleiotropic Genes and Pathways for the Comorbidity of Schizophrenia and Cardiometabolic Disease. Front Psychiatry 2020;11:256. [PubMed: 32425817]

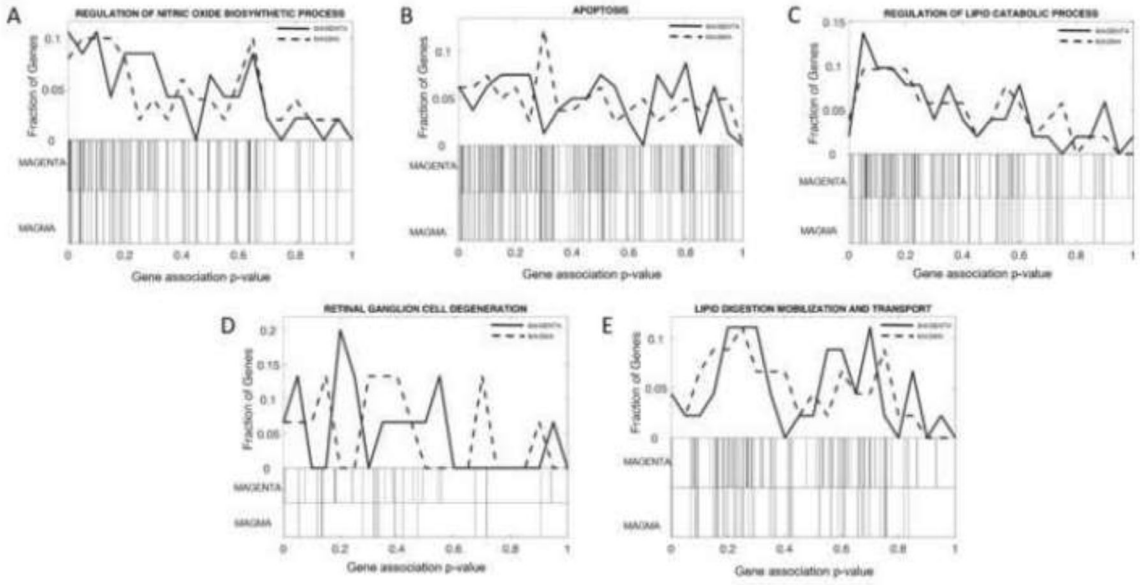

Figure 1. Distribution of DR gene association P-values for top 5 pathways.

The noncumulative distribution of confounder-adjusted gene association P values computed either with MAGENTA (solid line) or MAGMA (dashed line) is shown for five gene sets that passed multiple hypothesis correction (Benjamini-Hochberg) with one GSEA method and was nominally significant with the other method. The vertical lines in the two tracks mark the locations of the individual gene P-values based on MAGENTA (top track) or MAGMA (bottom track).

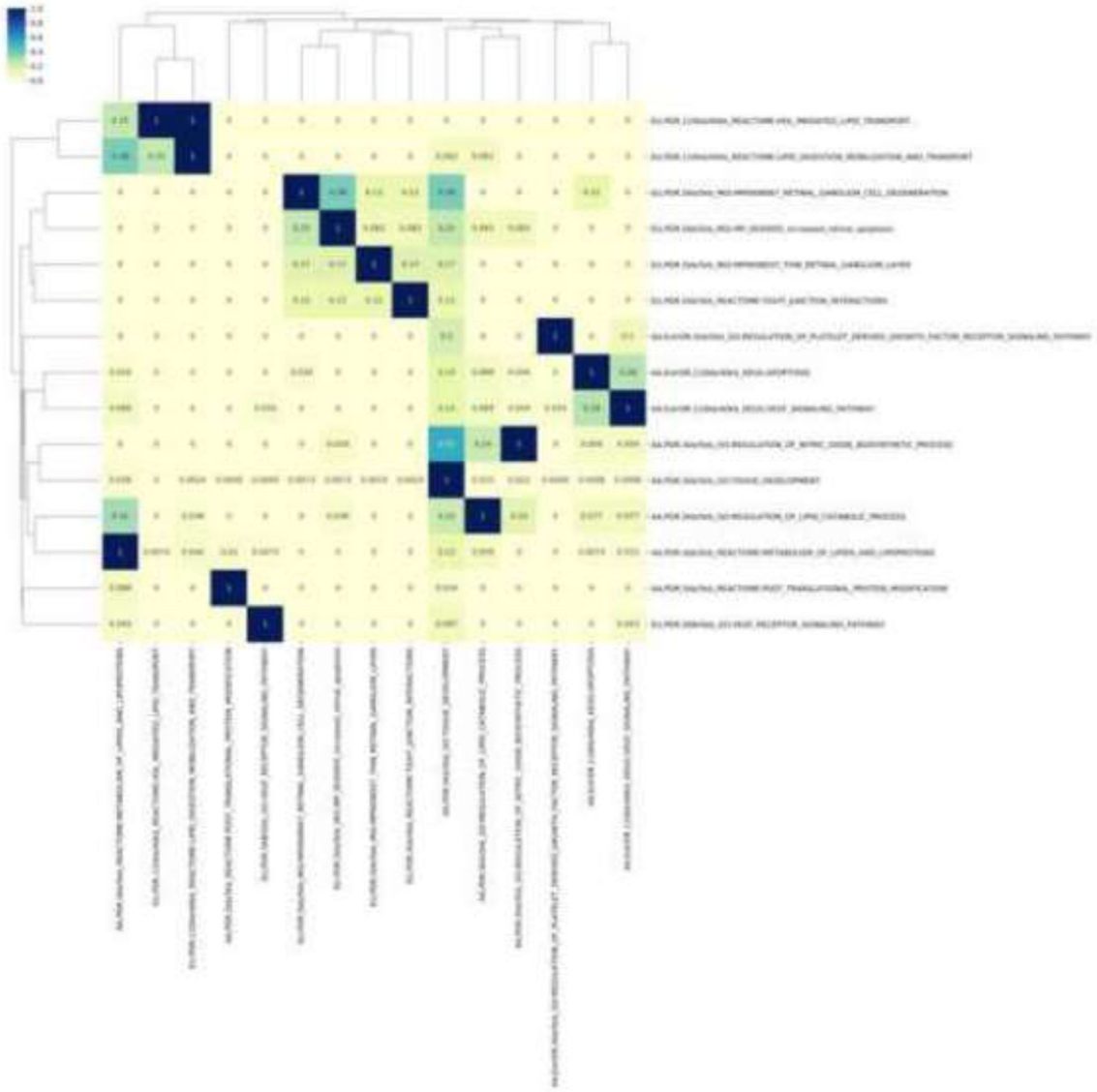

Figure 2. Heatmap of 15 significant gene sets clustered based on leading edge genes.

Gene sets were clustered based on overlap of their leading edge genes. The leading edge genes were determined based on the GWAS and gene boundaries provided in the gene set labels shown in the heatmap, using MAGENTA. The colorbar represents the fraction of leading edge genes in a given gene set in the row that overlap with the leading edge genes in the gene set in the column. The fractions are listed in the corresponding cell in the heatmap. AA = African American, EU = European, PDR = Proliferative Diabetic Retinopathy, DR = Diabetic Retinopathy.

Table 1.

Studies included in the gene set enrichment analysis

Study	Population	# of Eyes/# of Fields/Size of Fields Photographed	Glycemic Control Measure	Cases (ETDRS ≥ 14)	Ctrls (ETDRS < 14)	Cases (ETDRS ≥ 60)	Ctrls (ETDRS < 60)	Cases (ETDRS ≥ 30)
AAPDR	AA	2/7/30 deg.	HbA_{1C}	274	56	255	75	261
AGES*	EUR	2/2/45 deg.	HbA_{1C}	85	222	3	304	8
ARIC	AA	1/1/45 deg.	HbA_{1C}	96	265	3	358	73
	EUR	1/1/45 deg.	HbA_{1C}	126	632	6	752	80
AUST	EUR	NA‡	HbA_{1C}	522	435	187	770	346
BMES	EUR	2/5/30 deg.	FPG	124	208	1	331	37
CHS	AA	1/1/45 deg.	FPG	19	35	4	50	14
	EUR	1/1/45 deg.	FPG	26	119	4	141	16
FIND-Eye*	AA	2/2/45 deg.†	HbA_{1C}	330	167	264	233	303
	EUR	2/2/45 deg.†	HbA_{1C}	158	154	115	197	145
JHS	AA	2/7/30 deg.	HbA_{1C}	91	160	12	239	57
MESA	AA	2/2/45 deg.	HbA_{1C}	101	258	11	348	60
	EUR	2/2/45 deg.	HbA_{1C}	38	200	2	236	12
RISE/ RIDE	EUR	2/7/30 deg.	HbA_{1C}	--	--	80	117	--
WFU	AA	NA‡	HbA_{1C}	--	--	548	211	--
TOTAL	AA	--	Varies	911	941	1097	1514	768
TOTAL	EUR	--	Varies	1079	1970	398	2848	644

Ctrls= Controls, AAPDR = African American Proliferative Diabetic Retinopathy Study, AGES = Age, Gene/Environment Susceptibility Study, ARIC = Atherosclerosis Risk In Communities Study, AUST= Australian Genetics of Diabetic Retinopathy Study, BMES = Blue Mountains Eye Study, CHS=Cardiovascular Health Study, FIND-Eye = Family Study of Nephropathy and Diabetes-Eye, JHS = Jackson Heart Study, MESA = Multiethnic Study of Atherosclerosis, RIDE/RISE= Ranibizumab Injection in Subjects with Clinically Significant Macular Edema with Center Involvement Secondary to Diabetes, WFU=Wake Forest University, AA=African American, EUR = European, Illum=Illumina, Affy=Affymetrix, NA=not available, HbA_{1C}=hemoglobin A_{1C}, FPG=fasting plasma glucose, deg.= degrees, SNPs= single nucleotide polymorphisms, QC=quality control

†Not all FIND-Eye subjects had photographs but all participants had harmonization of exam and clinical data to an ETDRS score.

‡The AUST study used examination by an ophthalmologist to ascertain diabetic retinopathy.

The WFU study used a questionnaire to ascertain diabetic retinopathy.

Table 2.

Gene sets that were significant in either or both MAGENTA and MAGMA gene set enrichment analyses

Passed multiple hypothesis correction with MAGENTA			MAGENTA						MAGMA				
Population, GWAS, Gene Window Upstream/ Downstream	**Database: Gene set**	**Initial Gene Set Size**	**Effective Gene Set Size**	**Leading Edge Genes†**	**Proposed Number of Disease Associated Genes§**	**Fold-Enrichment**	**Uncorrected P value**	**Benjamini-Hochberg Corrected P value**	**Effective Gene set Size**	**Beta**	**Beta Std**	**Uncorrected P value**	**Benjamini-Hochberg Corrected P value**
AA, PDR Analysis, 5kb/5kb	**GO: REGULATION OF LIPID CATABOLIC PROCESS**	**52**	**51**	**26**	**13**	**2**	**0.0001***	**0.014**	**52**	**0.2051**	**0.0110**	**0.034 5**	**0.823**
AA, Extremes of DR Analysis, 110kb/40kb	KEGG: VEGF SIGNALING PATHWAY	76	70	29	11	1.61	0.0014*	0.018	75	0.0827	0.0053	0.2464	1.0
AA, PDR Analysis, 5kb/5kb	**GO: REGULATION OF NITRIC OXIDE BIOSYNTHETIC PROCESS**	**53**	**46**	**23**	**11**	**1.92**	**0.0003***	**0.022**	**50**	**0.3345**	**0.0175**	**0.0026**	**0.365**
AA, PDR Analysis,5kb/5kb	REACTOME: METABOLISM OF LIPIDS AND LIPOPROTEINS	478	448	135	23	1.21	0.0070	0.035	452	0.0185	0.0029	0.3121	1.0
AA, PDR Analysis, 5kb/5kb	GO: TISSUE DEVELOPMENT	1518	1444	410	49	1.14	0.0008	0.038	1457	0.0147	0.0040	0.2557	1.0
AA, PDR Analysis, 5kb/5kb	REACTOME: POST TRANSLATIONAL PROTEIN MODIFICATION	188	172	59	16	1.37	0.0040*	0.040	173	0.0751	0.0073	0.0942	0.942
AA, Extremes of DR Analysis, 110kb/40kb	**KEGG: APOPTOSIS**	**88**	**75**	**29**	**10**	**1.53**	**0.0063**	**0.041**	**81**	**0.1826**	**0.0122**	**0.0490**	**0.638**
AA, Extremes of DR Analysis, 5kb/5kb	GO: REGULATION OF PLATELET DERIVED GROWTH FACTOR RECEPTOR SIGNALING PATHWAY	14	14	10	6	2.5	0.0003*	0.043	14	0.2568	0.0071	0.1044	0.67
Passed multiple hypothesis correction with MAGMA			**MAGENTA**						**MAGMA**				
Population, GWAS, Gene Window Upstream/ Downstream	**Database: Gene set**	**Initial Gene Set Size**	**Effective Gene Set Size**	**Number of Genes Above Enrichment Cutoff**	**Proposed Number of Disease Associated Genes**	**Fold-Enrichment**	**Uncorrected P value**	**Benjamini-Hochberg Corrected P value**	**Effective Gene set Size**	**Beta**	**Beta Std**	**Uncorrected P value**	**Benjamini-Hochberg Corrected P value**
EU, PDR Analysis, 5kb/5kb	MGI: MP0030005 increased retinal apoptosis	36	35	12	3	1.33	0.1420	0.448	35	0.5760	0.0252	0.00001*	0.0005
EU, PDR Analysis, 5kb/5kb	REACTOME: TIGHT JUNCTION INTERACTIONS	29	27	8	1	1.14	0.3590	0.599	28	0.5808	0.0227	0.0001*	0.001

Passed multiple hypothesis correction with MAGENTA			MAGENTA						MAGMA				
Population, GWAS, Gene Window Upstream/ Downstream	**Database: Gene set**	**Initial Gene Set Size**	**Effective Gene Set Size**	**Leading Edge Genes†**	**Proposed Number of Disease Associated Genes§**	**Fold-Enrichment**	**Uncorrected P value**	**Benjamini-Hochberg Corrected P value**	**Effective Gene set Size**	**Beta**	**Beta Std**	**Uncorrected P value**	**Benjamini-Hochberg Corrected P value**
EU, PDR Analysis, 5kb/5kb	MGI: MP0008507 thin retinal ganglion layer	15	15	6	2	1.5	0.1430	0.419	15	0.6837	0.0196	0.0006*	0.012
EU, PDR Analysis, 110kb/ 40kb	REACTOME: HDL MEDIATED LIPID TRANSPORT	15	13	4	1	1.33	0.4020	1	15	0.6126	0.0175	0.0023*	0.023
EU, PDR Analysis, 110kb/ 40kb	**REACTOME: LIPID DIGESTION MOBILIZATION AND TRANSPORT**	**46**	**41**	**16**	**6**	**1.6**	**0.0306**	**0.306**	**45**	**0.3046**	**0.0150**	**0.0064**	**0.032**
EU, PDR Analysis, 5kb/5kb	GO: VASCULAR ENDOTHELIAL GROWTH FACTOR RECEPTOR SIGNALING PATHWAY	74	72	23	5	1.28	0.1070	0.958	72	0.3295	0.0207	0.0003*	0.045
EU, PDR Analysis, 5kb/5kb	**MGI: MP0008067 retinal ganglion cell degeneration**	**16**	**15**	**8**	**4**	**2**	**0.0175**	**0.359**	**15**	**0.5462**	**0.0157**	**0.0048**	**0.049**

AA = African American, EU = European, PDR = Proliferative Diabetic Retinopathy, DR = Diabetic Retinopathy, GO = Gene Ontology, KEGG = Kyoto Encyclopedia of Genes and Genomes, REACTOME = Reactome Pathway Database, MGI = Mouse Genome Informatics (Mouse Phenotype Ontology gene sets).

*These genes sets passed Bonferroni correction in the analysis (MAGENTA or MAGMA).

†Number of genes above 75th percentile enrichment cutoff.

§Number of observed genes above enrichment cutoff minus number of expected genes above enrichment cutoff.

The Uncorrected P value for MAGENTA was taken from 'NOMINAL_GSEA_PVAL_95PERC_CUTOFF' in Supplementary Table 2.

Gene sets in bold pass Benjamini-Hochberg with one method and are nominal significant (uncorrected P value <0.05) with the other method.

Table 3.

Correlation between gene association P-values computed by MAGENTA and MAGMA for top five gene sets.

Population, GWAS, Gene Window Upstream/ Downstream	Gene set database: Gene set Name	Gene set size	Pearson Correlation Coefficient	Pearson p-value	Spearman' s Rank Correlation Coefficient	Spearman p-value
AA, PDR Analysis, 5kb/5kb	GO: REGULATION OF NITRIC OXIDE BIOSYNTHETIC PROCESS	53	0.90	8.99×10^{-18}	0.93	7.49×10^{-21}
AA, PDR Analysis, 5kb/5kb	GO: REGULATION OF LIPID CATABOLIC PROCESS	52	0.86	4.78×10^{-16}	0.86	3.03×10^{-16}
AA, Extremes of DR Analysis, 110kb/40kb	KEGG: APOPTOSIS	88	0.95	4.25×10^{-42}	0.96	5.01×10^{-44}
EU, PDR Analysis, 110kb/40kb	REACTOME: LIPID DIGESTION MOBILIZATION AND TRANSPORT	46	0.86	6.28×10^{-14}	0.86	6.23×10^{-14}
EU, PDR Analysis, 5kb/5kb	MGI: MP0008067 RETINAL GANGLION CELL DEGENERATION	16	0.93	7.35×10^{-7}	0.92	7.99×10^{-7}

AA = African American, EU = European, PDR = Proliferative Diabetic Retinopathy, DR = Diabetic Retinopathy